Lab 6.13

Part 1

Steps taken:

1. Imported pandas, seaborn, numpy, and matplotlib libraries and turned csv file ‘product’ into a dataframe.
2. Turned the column ‘current\_retail\_price’ into a float.
3. Found the margin for the products by subtracting the current\_wholesale\_price from current\_retail\_price and added this new column to my table. Then sorted the table based upon the new margin column.

**Question 1**. The top 5 products that offer the best margins are: product id 81 – merchandise, product id 80 – merchandise, product id 83 – merchandise, product id 7 – whole bean/teas, and product id 81 – merchandise

1. Loaded csv file generations into a dataframe then counted the number of occurrences of each unique object.
   1. The amount of customers within each generation: gen z=21, baby boomers=19, gen x=15, older millennials=10, and younger millennials=5
2. Loaded csv files customer, and 201904 sales receipts, then joined these dataframes with the generations dataframe.
3. Once the three files were joined, calculated the profits by generations by multiplying the quantity by the unit price for the sales within each unique generation.

**Question 2.** Baby boomers spent the most on our products and are the most prevalent group of the customer base.

1. Opened the jupyter notebook file that was finally shared with us relating to the lab and copied and answered the questions that are relevant to answering the third question in the rubric.
2. Merged the dataframes containing information regarding the product categories and sales records. Then concatenated the columns for date and time to create a unique id for each transaction.
3. Grouped the sales by product to find the total quantity of each for all transactions.

**Questions 3.** The top 5 products that sell the most often are: Earl Gray Rg, Dark chocolate Lg, Latte, Morning Sunrise Chie Rg, and Ethiopia Rg.

1. Reordered by dataframe containing the transactions and products to show which products provide the highest profit margin.

**Question 4.** The products with the highest profit margins are merchandise, but the products which brought in the most profits was beverages. Produced 2 bar charts to show this.

1. Separated loyalty customers from non-loyalty customers and made a pie graph to show the distribution of members to non-members.
2. Calculated the total profits from each group. Then made bar charts representing the profits. They are too similar to gather a clear conclusion of a strategy.
3. Separated transactions by whether or not they were in store or out of store and made a pie chart to see the difference. They too were too similar to gather a clear conclusion of a strategy.

**Question 5/6.** The segments of customers evaluated above show little variation and thus makes forming a marketing strategy difficult. Based upon the information relating to generational differences in spending, it would be wise to consider catering to the older generations. However, we likely also want to expand our customer base by marketing to the younger generations and encourage them to sign up and participate in the loyalty program so we can continue to gather information such as this.

Data Limitations:

1. The data we have relating to the customer base is only actually related to the customers who are part of the loyalty program and not the casual customers. In addition, if a loyalty program member does not provide their loyalty ID, then the sale will not be attributable to that customer and may skew our results even further.
2. Along with this, the file with all the transactions does not contain the customer's loyalty number vs those who do not have a number leaving the program to auto fill it with zero instead of a null value, making it more difficult to separate the two groups and make an analysis based on this.
3. Some of the data gathered is unclear as to what it means. For example: line\_item\_amount appears as though it is the product of the quantity and unit price, but this is not made clear with a data dictionary. Also, one of the data points is instore\_yn. This, I assume, means that the item was paid for either in the store or out of store, perhaps on an app or maybe just over the phone. This is not specified as to what this really means and what and out of store purchase would be.
4. The process of becoming a loyalty program member is unclear as well. If there were inconvenient obstacles in the way of entering the program, the younger generations may be less willing to sign up. The more technology spreads, the more businesses have to cater to those who utilize it heavily.

Insights:

This data has a lot of information in it as evident by the many different csv files and this is very helpful, if not essential, for making a detailed data analysis. The data limitations make it clear that even if you have access to a lot of data, there is always room for more. Even though people are seemingly more and more turned off by the idea of a random business gathering a lot of detailed information about their customers, it’s obvious that it allows for further insight and thus catering to the behaviors of customers both now and in the future.

This particular simulated business could benefit from gathering as much data possible so as to cater their marketing campaign towards those who have not entered into the loyalty program and what may be in the way of that happening. Perhaps the customer base is mostly older generations just because they have kept their loyalty active the longest. Or perhaps the younger generations have little interest in the business for some reason unbeknownst to us. It’s hard to say for sure, but more data, perhaps comments even, would be beneficial.

Actions:

* Take comments relating to the ease of signing up for the loyalty program and try to address them.
* Market towards all generations and encourage signing up for the loyalty program whenever possible.
* There could be room for more marketing of the products with the highest profit margins – merchandise – in order to increase revenue. Possibly even considering specially modified made-to-order products.
* Knowing which products sell the most will help to curb excess inventory by not stocking much more than what usually sells.